AP Statistics Review
Unit 2 Quiz (Chapters 7 and 8)

Name: _____

Part I: Multiple Choice
Circle the best answer.

1. Paired data for the percentage of men and women who smoke cigarettes is recorded for selected years from 1965 to 2002. The variable Male stands for the percentage of men who smoke in a given year; Female stands for the percentage of women who smoke in a given year. Summary statistics for these measurements are given below. Which of the following is the equation of the Least-Squared Regression Line for Females vs. Males?

Descriptive Statistics: Male, Female

Variable	Mean	Std. Dev.
Male	30.38	6.89
Female	24.884	4.013

Correlation of Male and Female = 0.950

   a. $\hat{Female} = 8.07 + 0.5533 Male$
   b. $\hat{Female} = 16.61 + 0.5533 Male$
   c. $\hat{Female} = 8.91 + 0.5257 Male$
   d. $\hat{Female} = -10.21 + 1.6311 Male$
   e. $\hat{Female} = -24.67 + 1.6311 Male$

2. A an educational researcher found that there was a strong positive relationship between the age a child learns to read (AR—in months) and the age the child is able to add single digit numbers (AA—in months). The Least-Squares Regression Line that the researcher conducted for her Linear Model was: $\hat{AA} = 3.5 + 3.6 AR$
Which of the following is the best interpretation of the slope of the regression line?
   a. An increase in reading age of 1 month is associated with an increase of about 3.5 months in the adding single digit number age.
   b. An increase in adding single digit number age of 1 month is associated with an increase of about 3.5 months in the reading age.
   c. An increase in the reading age of 1 month is associated with an increase of about 3.6 months in the adding single digit number age.
   d. An increase in the adding single digit number age is associated with an increase of about 3.6 months in the reading age.
   e. None of the above interpret the slope correctly.

3. A sports researcher is studying statistics for a set of professional baseball teams over a single season. The researcher performs a regression analysis on the total payroll (in millions of dollars) versus the number of wins. An output of the results is shown below. Which of the following is the best interpretation for the slope of the regression line?

Regression Analysis: Wins versus Payroll

Predictor	Coef	SE	T	P
Constant	66.42	12.46	5.33	0.000
Payroll	0.2070	0.1673	1.24	0.240

S = 16.9082 R-Sq = 11.3% R-Sq(adj) = 3.9%

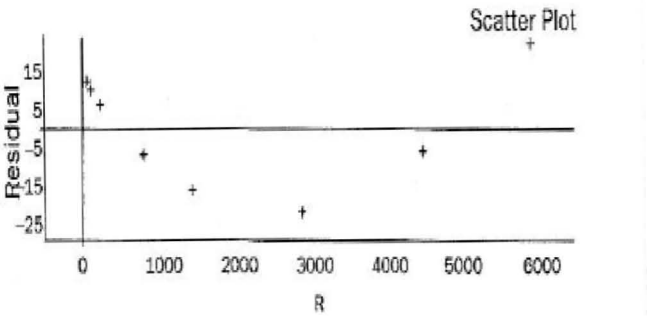
   a. An increase in payroll of $5 million is associated with an increase of about 1 win, on average.
   b. An increase in payroll of $5 million is associated with a decrease of 1 extra win, on average.
   c. An increase in payroll of $1 million is associated with a decrease of about 1 win, on average.
   d. An increase in payroll of $1 million is associated with an increase of about 2 wins.
   e. None of the above interprets the slope correctly

4. Using the Regression Analysis in #3, which of the below is the best interpretation of the R^2 Value?
   a. 11.3% of the variation in payroll can be accounted for by the variation in the number of wins, on average.
   b. 11.3% of the variation in wins can be accounted for by the variation in payroll, on average.
   c. The LSRL contains 11.3% of the data points in the scatterplot.
   d. The LSRL correctly predicts the payroll 11.3% of the time..
   e. The LSRL correctly predicts the number of wins 11.3% of the time.

5. The correlation between two scores X and Y is -0.71. Which of the following is a correct statement about the relationship between the values of X and Y?
   a. 71% of the variation in Y is explained by the least-squares regression of Y on X.
   b. There is a moderate tendency for an increase in the value of X to cause a decrease in the value of Y.
   c. There is a moderate tendency for an increase in the value of X to cause an increase in the value of Y.
   d. Among the values of X and Y, there is a moderate tendency for decreasing values of X to correspond to decreasing values of Y.
   e. Among the values of X and Y, there is a moderate tendency for decreasing values of X to correspond to increasing values of Y.

6. A study was conducted that analyzed the relationship between the height of teenage males and their weight. The relationship was linear and had an r-value of 0.875. The LSRL was $\hat{weight} = 40 + 1.5height$. Height is expressed in inches and weight in pounds. If a female's height is 1.25 standard deviations above the mean, how many standard deviations from mean would we predict is the weight of that female?
    a. 1.5
    b. 0
    c. -1.5
    d. 1.09
    e. -1.09

7. For the nine planets in the solar system, a list is made of the lengths of a year Y (measured in earth years) and the mean orbital radius *R* (measured in millions of km). A regression analysis yields $\hat{Y} = -13 + .04.8R,\ r^2 = 0.98,$ and the following residual plot. Which of the following is the most accurate statement about the regression results?



    a. The value of r^2 shows there is a strong linear relationship between the length of year and mean orbital orbit.
    b. The residual plot shows there is a strong linear relationship between length of year and mean orbital orbit.
    c. The slope of the regression line shows there is a negative association between length of year and mean orbital radius.
    d. The regression line underpredicts the length of a year for a planet with mean orbital radius of 100 million km.
    e. All of the above are accurate statements

8. A calculus teacher is investigating the relationship between his students' grades prior to their final exam (PRIOR) and their grade on the final exam itself (EXAM). He conducts a regression analysis of his students from last year with the following results. According to the regression line, which of the following is the predicted final exam grade for a student who has an average of 85 prior to the final?

Source	DF	Sum of Sq	Mean Squ	F-ratio
Regression	1	2290.3	2290.3	38.52
Residual Error	30	1783.5	59.5	

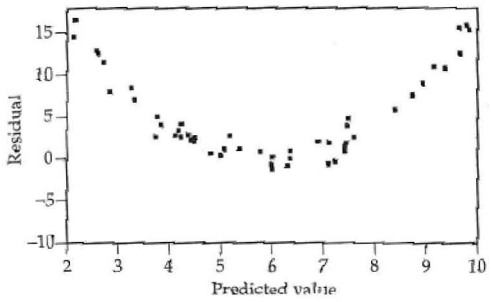
Predictor	Coef	SE Coef	T	P
Constant	8.63	10.08	0.86	0.399
PRIOR	0.8126	0.1309	6.21	0.000

S = 7.71046 R-Sq = 56.2% R-Sq(adj) = 54.8%

    a. 21
    b. 48
    c. 78
    d. 79
    e. 85

9. Data was collected on two variables *X* and *Y* and a least squares regression line was fitted to the data. The estimated equation for this data is $\hat{y} = -2.29 + 1.70x$. What is the residual for the point (5,6)?
    a. 7.91
    b. 6.21
    c. 0.21
    d. -0.21
    e. -2.91

10. Sixty pairs of measurements were taken at random to estimate the relation between variables *X* and *Y*. A Least-Squares Regression Line was fitted to the collected data. The resulting residual plot is as follows. Which of the following conclusions is appropriate?



    a. A line is an appropriate model to describe the relation between *X* and *Y.*
    b. A line is not an appropriate model to describe the relation between *X* and *Y.*
    c. The assumption of normality of errors has been violated.
    d. The assumption of constant sample standard deviations has been violated.
    e. The variables *X* and *Y* are not related at all.

11. The computer output from the regression equation of a study done relating the number of baggage handlers and the number of arriving passengers is shown below. What is the value of the correlation coefficient for the number of baggage handlers and number of arriving passengers?
    a. -0.842
    b. 0.651
    c. 0.709
    d. 0.842
    e. 1.562

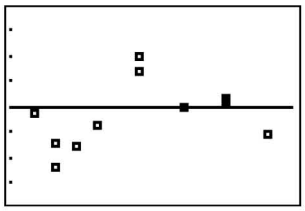
Predicted Baggage Handlers = 2.86 + 0.00408 (number of passengers)

Predictor	Coef	StDev	T	P
Constant	2.860	1.324	2.16	0.083
Passengers	0.004081	0.001168	3.49	0.017
S = 1.562	R-sq = 70.9%		R-Sq(adj) = 65.1%	

12. Circle the choices are true statements.
    a. An r-value of 0.043 means that there is no relationship between the two quantitative variables.
    b. In the LSRL, the sum of the residuals is always zero.
    c. A positive residual means that the LSRL overestimated the predicted value.
    d. If you have an R^2-value of 98%, it means that your linear model is very accurate.
    e. If $\hat{y} = 9.9 + 4x$ and the R^2-value is 81%, then r=.9
    f. The r-value tells you the strength and direction of a linear relationship.

g. In order to calculate a Least-Squares Regression Line, you need to satisfy the Nearly Normal Condition.

h. Suppose that I calculated a LSRL line of $\hat{y} = 9.9 + 4x$ and then created the residual plot below. This residual plot tells me that my linear model in not appropriate.



i. A residual plot helps to satisfy the Straight Enough Condition.

j. A residual value of zero means that the predicted value and the observed value are equal.

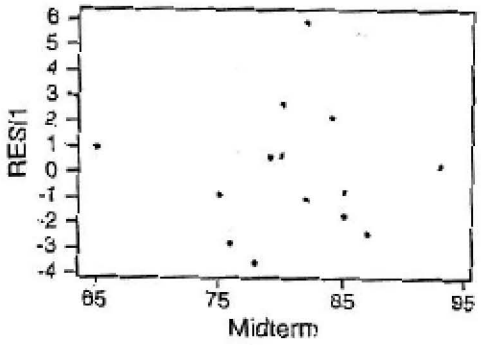
Part II: Free-Response.

13. The graph below is the residual graph created from the regression analysis of midterm exam scores vs. final exam scores for an AP Statistics class. The explanatory variable was midterm exam scores and the response variable was final exam scores. The regression equation is $\hat{Final} = 11.7 + 0.882 Midterm$.

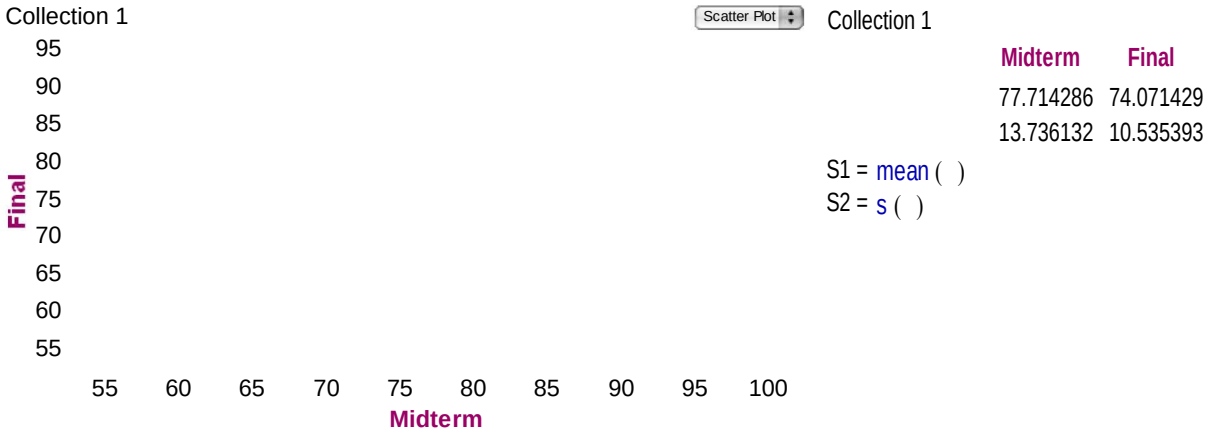
a. What was the approximate final exam score for the student whose midterm score was 65?

b. There were three scores that had a residual value of zero. What does this mean?

c. About how many scores were underpredicted by the regression line? Circle those scores on the residual plot. Also, explain why those scores are underpredicted by the regression line.



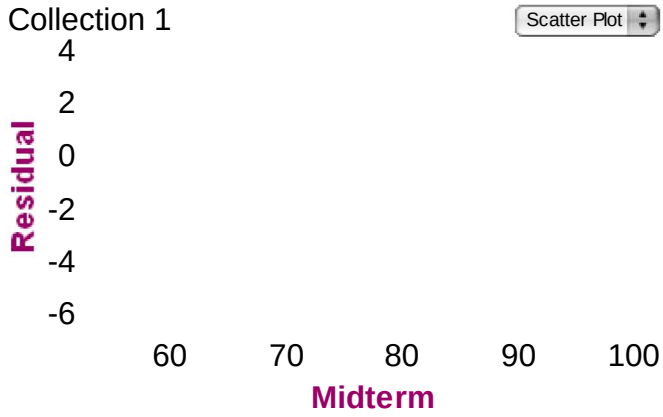
14. The graph below is the scatterplot showing the relationship between the midterm exam scores and final exam scores of a college economics course. The residual plot of the data is shown below the graph, as is the least squared regression line and the R-squared value. Next to the graph is a table summarizing the mean and standard deviation of both variables.



Collection 1

	Midterm	Final
	77.714286	74.071429
	13.736132	10.535393

S1 = mean ( )
S2 = s ( )



LSRL: $\hat{Final} = 17.318 + 0.7303 Midterm$ and $R^2 = 91\%$

a. Describe the association between Final exam score and Midterm exam score.
b. Explain in context what the slope means.
c. Explain in context what the *y*-intercept means.
d. Explain in context what R^2 means.
e. Is the model appropriate for the data given? Explain.
f. In this context, what does a negative residual mean?
g. What if, for some strange reason, you decided to switch the variables, making the final exam scores the predictor variable and the midterm the explanatory variable. What would be the slope of the LSRL for that scatterplot. Show any necessary work and assume that the model is appropriate for the data.
h. Suppose your midterm exam was .8 Standard Deviations below the mean. Predict how many standard deviations from the mean your final exam score would be?